The Prediction of Shipping Traffic Flow Based on GA-SVM Model

Jiapei Deng1,2, Yadong Yang1,2
1. School of Navigation, Wuhan University of Technology, Wuhan, Hubei China;
2. Hubei Key Laboratory of Inland Navigation Technology, Wuhan, Hubei China

ABSTRACT

E-dong international airport will be built as international freight airport in Hubei province, which is for promoting economic development. E-dong international airport is near the Yangtze River, now we have to ensure the safety of aircraft and vessel, so research on prediction of shipping traffic flow is necessary, because it can provide basic date for the navigation safety management and play an important role in reducing traffic accident. As we all know that the key problem of Support vector regression is to solve large-scale training sample, So in order to forecasting shipping traffic flow accurately, we adopt SVR prediction model optimized by genetic algorithm. Then the shipping data which we collected from Huangshi Maritime Management has been adopted to verify these models presented in the paper. Comparing with the results of BPNN and SVR, which showed that GA-SVR model is more accurate in predicting ship traffic flow.

KEY WORDS: Shipping Traffic Flow; BPNN; Support Vector Regression; Genetic Algorithm; GA-SVR.

INTRODUCTION

In order to build the international freight airport, Hubei Provincial Government and the SF Group signed the cooperation agreement on Hubei international freight airport project in December 2017. They decided to establish airport in Yanji of Ezhou. The airport is near the Yangtze River channel, before we research that the shipping traffic flow of nearby waters is how to effect aircraft, we should understand the change rule of the shipping traffic flow first. So we did research on prediction of the shipping traffic flow of passing the section of E-dong Yangtze River Bridge in this paper.

At present, there are several methods for forecasting the shipping traffic flow at home and abroad, such as regression analysis and grey relational analysis (Feng HG et all, 2011; Li YC et all, 2004). However, Feng, K et all (2015) discovered that these prediction methods are difficult to satisfy the prediction efficiency and accuracy. Because training speed of the existing prediction methods is slower and prediction accuracy rate is not higher (Niu YL, 2017), this paper adopts a new model for predicting the ship traffic flow, named GA-SVR prediction model, which is used for shipping traffic flow prediction firstly. (Srinivas M, 1994) find that Genetic algorithm is based on the evolution rule of biological world, and is good at searching randomly. SVR has many unique advantages in solving nonlinear and high-dimensional pattern recognition (Venkoba Raoa,B, 2009). So we adopted the SVR model optimized by GA to forecast shipping traffic flow passing through the section of E-dong Yangtze River Bridge, which will provide the basis for the management of the ship of navigable waters near the airport.

SUPPORT VECTOR REGRESSION MODEL

The principle of Support Vector Regression

The basic idea of Support Vector Regression is to map the input sample data $x$ to high-dimensional feature space linear regression function $f(x)$ by using a nonlinear function $\phi(x)$ (Vladimir, 2004; Suykens J A K, 2001), so as to transform the nonlinear regression problem into the linear problem of high-dimensional space. By using the principle of structural risk minimization, we introduce the concept of interval and kernel function (Cristianini N, 1999), and then use the kernel function of original space to replace the inner product operation in high-dimensional space for avoiding complex computation.

The linear regression function according to the theory of statistics is:

$$f(x) = w \cdot \phi(x) + b \quad w, x \in \mathbb{R}^n, b \in \mathbb{R}$$  \hspace{1cm} (1)

in Eq.1, the $w$ is the weight vector and the $b$ is the offset.

According to the theory of statistics, the nonnegative slack variable is introduced, then the SVR problem is changed to:

$$f(x) = \min_{\xi_i \geq 0} \left[ \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \right]$$  \hspace{1cm} (2)

$$\text{s.t. } \begin{cases} y - w \cdot \phi(x) - b \leq \varepsilon + \xi_i \\ w \cdot \phi(x) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

In Eq.2, $C$ is an error penalty parameter, and $\varepsilon$ is an insensitive loss function. Obviously, this is a convex two - time program, and its dual form can be obtained.

$$f(x) = \sum_{i=1}^{n} (a^i - a^i_*) k(x_i, x) + b$$  \hspace{1cm} (3)